

# 期末專案參考資訊



# 物件辨識

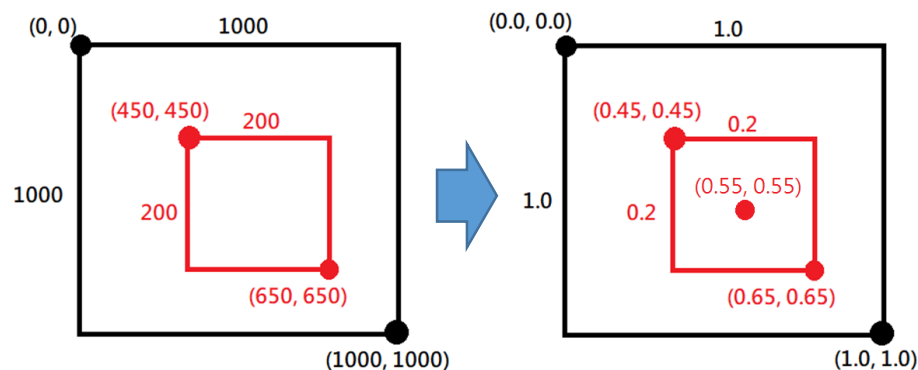


# 資料集座標讀取

- Kitti資料集中label的格式如下:

```
Misc 0.00 0 -1.82 804.79 167.34 995.43 327.94 1.63 1.48 2.37 3.23 1.59 8.55 -1.47
```

- 其中:
  - (804.79, 167.34, 995.43, 327.94)分別為( $X_{min}$ ,  $Y_{min}$ ,  $X_{max}$ ,  $Y_{max}$ )
- YOLO的label格式如下:
  - 1 0.716797 0.395833 0.216406 0.147222
  - 分別為<object-class> <x> <y> <width> <height>
  - <x> <y> <width> <height>分別是label的中心x、y座標與長寬，數值範圍為(0.0, 1.0]，此範圍是YOLO訓練所需格式，是針對影像長寬進行正規化後的結果。
  - 正規化就是假設圖片長寬為1的時候，label在這張圖片中的位置。
    - 範例:假設Kitti有長寬1000x1000的影像，其中有一個label的( $X_{min}$ ,  $Y_{min}$ ,  $X_{max}$ ,  $Y_{max}$ )座標為(450, 450, 650, 650)，則YOLO的 <x> <y> <width> <height>為 0.55 0.55 0.2 0.2





# 資料集座標讀取

$$x = (xmin + (xmax-xmin)/2) * 1.0 / image\_w$$

$$y = (ymin + (ymax-ymin)/2) * 1.0 / image\_h$$

$$w = (xmax-xmin) * 1.0 / image\_w$$

$$h = (ymax-ymin) * 1.0 / image\_h$$

類別 代碼	物件中心x位在整張圖片 x的比例	物件中心y位在整張圖片 y的比例	物件寬度w佔整張圖片寬 度的比例	物件長度h佔整張圖片長 度的比例
Categ ory  num ber	Object center  in X	Object center  in Y	Object width  in X	Object height  in Y



# 資料集座標讀取

- 因此可以自己寫程式，讀取txt檔，將 $(X_{min}, Y_{min}, X_{max}, Y_{max})$ 轉為`<object-class> <x> <y> <width> <height>`，這樣就不必手動標記了，並儲存成與圖片相同的檔名
  - `object-class` → 依自己label數量而定有n種，就有0~n-1，可以只label一種就好，方便計算iou、準確率等等
  - $x = f( (X_{min} + X_{max}) / 2 )$
  - $y = g( (Y_{min} + Y_{max}) / 2 )$
  - $width = f( X_{max} - X_{min} )$
  - $height = g( Y_{max} - Y_{min} )$
  - $f(x)$ 代表將x座標根據圖片寬度正規化到(0.0, 1.0]
  - $g(y)$ 代表將y座標根據圖片長度正規化到(0.0, 1.0]



# 計算準確度

- ground truth 就是你標記的座標(從kitty轉成YOLO的那個座標)。  
預測結果座標要和這個座標比對準確率。
- YOLO除了輸出辨識結果圖片之外也可以從code中取得預測框的座標，可以使用iou來算準確率。
  - 一個預測框跟所有ground truth的iou=0就是代表誤框(或沒框到)。
  - 一個預測框跟ground truth的iou可以設定0.5來決定是否為誤框(因為可能框不完整)。若有多個iou就以最大的iou為主。



# YOLO參數

- YOLO參數說明，請參考：
  - <https://chih-sheng-huang821.medium.com/%E6%B7%B1%E5%BA%A6%E5%AD%B8%E7%BF%92-%E7%89%A9%E4%BB%B6%E5%81%B5%E6%B8%ACyolov1-yolov2%E5%92%8Cyolov3-cfg-%E6%AA%94%E8%A7%A3%E8%AE%80-75793cd61a01>
- YOLOv1~v4不同版本只是網路架構有稍微修改而已，其餘參數皆大同小異。



# 辨識多種類別

- 各個類別的訓練樣本數量要盡量保持一致，並且都要有一定的數量，。例如，欲同時辨識人、汽車、機車，則這三類樣本數量都要有差不多的數量，如1000:1000:1000。這樣訓練出來的模型用來辨識才會有鑑別度。





# 軌跡預測



# 軌跡預測

- 可以自行決定要使用Social LSTM, Social GAN等方法，請根據Lecture內容中提到的參數去調整來訓練。
- 軌跡預測，就說若是時間不足，可以用LSTM不一定要social LSTM，但social準確度會更高。
- 其中ADE(平均位移誤差 Average displacement error)與FDE(終點位移誤差 Final displacement error)，可以在進行validate後輸出的結果看到數值結果。



# 軌跡預測

- 軌跡預測訓練時若遇到錯誤，可先確認是否與資料集格式有關，期末專案資料集中的座標信息，與先前project實作時所使用的資料集可能有所差異，請自行調整至相同格式。



# Social GAN資料集格式

- 內建資料集格式:

- 檔案位置: SocialGAN2\datasets\04171525\train\new\_annotation.txt。
- 每一欄分別為 <FrameID> <objectID> <中心點x(影像寬度/100)> <中心點y(影像長度/100)> <objectType>

```
0      5      6.34    4.44    Sedan
0      6      5.92    2.27    MotorcycleWithRiderWithHelmet
0     22      6.48    2.47    BicycleWithRider
1      5      6.5     4.41    Sedan
```

- VIRAT資料集格式:

- README\_format\_release2.txt 中有針對各欄數值做說明。
- 以 VIRAT\_S\_000001.viratdata.objects.txt 為例:
- 每一欄分別為 <Object id> <Object duration> <Current frame> <bbox lefttop x> <bbox lefttop y> <bbox width> <bbox height> <Object Type>

```
1 385 3462 1 663 76 132 1
1 385 3463 5 663 76 132 1
1 385 3464 10 663 76 132 1
1 385 3465 15 663 76 132 1
```